



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/668,121	09/23/2003	Sherif Yacoub	200207195-1	1995

22879 7590 05/22/2008
HEWLETT PACKARD COMPANY
P O BOX 272400, 3404 E. HARMONY ROAD
INTELLECTUAL PROPERTY ADMINISTRATION
FORT COLLINS, CO 80527-2400

EXAMINER

SAINT CYR, LEONARD

ART UNIT	PAPER NUMBER
----------	--------------

2626

NOTIFICATION DATE	DELIVERY MODE
-------------------	---------------

05/22/2008

ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

JERRY.SHORMA@HP.COM
mkraft@hp.com
ipa.mail@hp.com

DETAILED ACTION

Response to Arguments

1. Applicant's arguments filed 02/21/08 have been fully considered but they are not persuasive.

Applicant argues that neither Bennett et al., nor Murveit et al., teach or suggest analyzing characteristics of the first speech utterance to determine an age and a gender of the first user (Amendment, pages 7 - 11).

The examiner disagrees, Bennett et al., teach that “a more complex characteristic of the incoming stream is contextual information. Contextual information is that information related to the environment around the input stream. Contextual information may include gender, age, ethnicity” (paragraph 18; paragraph 31, line 3). Using contextual information related to the environment around the input stream to determine gender, age and ethnicity implies analyzing characteristics of the first speech utterance to determine an age and a gender of the first user.

Applicant argues that neither Bennett et al., nor Murveit et al., teach or suggest a ranking matrix for selecting one of the recognizers (Amendment, pages 7 – 11).

The examiner disagrees, Bennett et al., teach that “if the system knows that the user is dictating a legal memo based on the current state of the dialog, it may use the legal-dictation-optimized recognizer” (paragraph 33, lines 19 – 21). Choosing the legal-

dictation-optimized recognizer among of optimized for legal use, optimized for medical use, and for general use implies a ranking matrix for selecting one of the recognizers.

Claim Rejections - 35 USC § 103

2. The text of those sections of Title 35, U.S. Code not included in this action can be found in a prior Office action.

3. Claims 1 – 20 rejected under 35 U.S.C. 103(a) as being unpatentable over Bennett et al., (US PAP 2002/0194000), in view of Murveit et al., (US Patent 7,058,573)

As per claims 1, 8, and 14, Bennett et al., teach an automatic speech recognition (ASR) that comprises:

providing a plurality of categories (“American male”) for different speech utterances; assigning a different ASR engine to each category to develop a ranking matrix (“recognizers that have good performance for American men southern accents be enabled”) based on the ranks of the ASR engines (“select the best recognizer and its results”; paragraph 15, lines 6 – 9; paragraph 19; paragraph 20, lines 7 – 9; Abstract, lines 7, and 8);

processing the different speech utterances at different ASR engines (“the speech recognition system enable some of the speech recognizers and received results”; abstract, lines 4 – 6)

receiving a first speech utterance (“receiving the input stream”) from a first user; (paragraph 12, lines 1, and 2; paragraph 19, lines 10 - 12);

analyzing characteristics of the first speech utterance to determine an age and a gender of the first user; classifying the first speech utterance into one of the categories based on the age and gender of the first user (“a more complex characteristic of the incoming stream is contextual information. Contextual information is that information related to the environment around the input stream. Contextual information may include gender, age, ethnicity”; paragraph 18; paragraph 31, line 3);

extracting characteristics about the first user from content of the first speech utterance to classify the first speech utterance into one of the categories; and consulting the ranking matrix to select a single one of the ASR engines assigned to the category to which the first speech utterance is classified to automatically recognize the first speech utterance (“a user calls into the system and navigates the menus using control keywords and then starts a dictation process. Additionally, a variety of recognizers are optimized for dictation may be available, for example. If the system knows that the user is dictating a legal memo based on the current state of the dialog, it may use the legal-dictation optimized recognizer”; paragraph 33, lines 8 – 21).

However, Bennett et al., do not specifically teach receiving ground truths with correct text for the different speech utterances; and comparing output from the each of the different ASR engines with the ground truths to determine ranks of the different ASR engines for accuracy in recognizing the different speech utterances.

Murveit et al., teach assuming the spoken input is the word, “Boston”. The assigned score is a probability or is related to the probability that the corresponding expression correctly corresponds to the spoken input. The expression with the highest

assigned score or certainty is selected as the output (probability that the corresponding expression correctly corresponds to the spoken input implies comparing output from the each of the different ASR engines with the ground truths to determine ranks of the different ASR engines for accuracy in recognizing the different speech utterances, since the highest score is selected among all the assigned scores; col.2, lines 56, and 57; col.5, lines 21 – 23; col.9, lines 22 – 24).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to assign scores based on expression correctly corresponds to the speech input as taught by Murveit et al., in Bennett et al., because that would maintain a high degree of recognition accuracy in a speech recognition system (col.2, lines 33, and 34).

As per claims 2, 9, 15, and 20, Bennett et al., further disclose the ranking matrix is a table that defines which ASR engine or combination of ASR engines has a best accuracy (“may use the legal-dictation-optimized recognizer”) for different ages and genders of users (paragraph 33, lines 19 – 21; paragraph 18; paragraph 31, line 3)

As per claim 3, Bennett et al., further disclose assigning a different ASR engine to each category further comprises assessing accuracy of each ASR engine for each category (“accuracy of each recognizer in a particular situation”; paragraph 22, lines 8, and 9).

As per claims 4, and 16, Bennett et al., further disclose assessing accuracy of each ASR engine for each category further comprises determining a least Word Error Rate of each ASR engine for each category (“a recognizer with a recognizer-based confidence value of 90%”; paragraph 42, lines 3, and 4).

As per claim 5, Bennett et al., further disclose assigning a different ASR engine to each category further comprises assessing time required for each ASR engine to recognize speech utterances (“performance over time”; paragraph 42, line – paragraph 43, line 3).

As per claim 6, Bennett et al., further disclose receiving a second speech utterance from a second user; classifying the second speech utterance into one of the categories; and selecting the ASR engine assigned to the category to which the second speech utterance is classified to automatically recognize the speech utterance, wherein the ASR engine assigned to the category to which the second speech utterance is classified is different from the ASR engine assigned to the category to which the first speech utterance is classified (using characteristics of the communication channel and contextual information such as gender to enable some of the recognizers among a plurality of recognizers, implies that it is inherent to classify another speech to another category; paragraph 20; paragraph 17; paragraph 31, line 3).

As per claim 7, Bennett et al., further disclose that the first speech utterance is classified into a male category, and the second speech utterance is classified into a female category (“gender”; paragraph 19, lines 10 – 12; paragraph 31, line 3).

As per claim 10, Bennett et al., further disclose different categories are selected from the group consisting of gender, noise level, and pitch (“signal strength”; paragraph 15, line 7; paragraph 31, line 3).

As per claim 11, Bennett et al., further disclose different ASR engines comprise single ASR engines (“single recognizer”) and multiple ASR engines combined together (paragraph 21, lines 1, and 2; paragraph 20, lines 7, and 8).

As per claim 12, Bennett et al., further disclose the plurality of different ASR engine rankings are derived from statistical analysis (“performance history of the particular recognizer”; paragraph 23, line 5).

As per claim 13, Bennett et al., further disclose that the statistical analysis comprises assessing accuracy of speech recognition of different ASR engines with different speech signals (“accuracy of each recognizer in a particular situation”; paragraph 22, lines 8, and 9).

As pre claim 17, Bennett et al., further disclose at least three different ASR engines and at least three different combination schemas of ASR engines to represent a total of at least six different ASR engines (“processing cell phone audio stream with some recognizers among multiple recognizers”; paragraph 10, lines 2, and 3; paragraph 16, lines 2 – 4).

As per claim 18, Bennett et al., further disclose that a telephone network comprising at least one switching service point coupled to the computer system (“output switch 16”; paragraph 4, lines 8 – 10; paragraph 10; paragraph 13, line 3).

As per claim 19, Bennett et al., further disclose that at least one communication device in communication with the switching service point to provide the speech utterance (“cell phone connection” paragraph 10; paragraph 13, line 3).

As per claim 10, Bennett et al., further disclose different categories are selected from the group consisting of gender, noise level, and pitch (“signal strength”; paragraph 15, line 7; paragraph 31, line 3).

As per claim 11, Bennett et al., further disclose different ASR engines comprise single ASR engines (“single recognizer”) and multiple ASR engines combined together (paragraph 21, lines 1, and 2; paragraph 20, lines 7, and 8).

As per claim 12, Bennett et al., further disclose the plurality of different ASR engine rankings are derived from statistical analysis (“performance history of the particular recognizer”; paragraph 23, line 5).

As per claim 13, Bennett et al., further disclose that the statistical analysis comprises assessing accuracy of speech recognition of different ASR engines with different speech signals (“accuracy of each recognizer in a particular situation”; paragraph 22, lines 8, and 9).

Conclusion

5. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire **THREE MONTHS** from the mailing date of this action. In the event a first reply is filed within **TWO MONTHS** of the mailing date of this final action and the advisory action is not mailed until after the end of the **THREE-MONTH** shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than **SIX MONTHS** from the mailing date of this final action.

Art Unit: 2626

6. Any inquiry concerning this communication or earlier communications from the examiner should be directed to LEONARD SAINT CYR whose telephone number is (571)272-4247. The examiner can normally be reached on Mon- Friday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571) 272-7602. The fax phone number for the organization where this application or proceeding is assigned is (571)-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

LS
05/13/08

/Richemond Dorvil/
Supervisory Patent Examiner, Art Unit 2626